Glycobiology vol. 14 no. 12 pp. 1229–1246, 2004
doi:10.1093/glycob/cwh106
Advance Access publication on June 22, 2004

Tracking global patterns of *N*-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin

Ming Zhang[2], Brian Gaschen[2], Wendy Blay[3,4], Brian Foley[2], Nancy Haigwood[3,4], Carla Kuiken[2], and Bette Korber[1,5]

[2]Theoretical Biology Group, Los Alamos National Laboratory, Los Alamos, NM 87544; [3]Seattle Biomedical Research Institute, Seattle, WA 98109; [4]Pathobiology Department, University of Washington, Seattle, WA 98195; and [5]The Santa Fe Institute, Santa Fe, NM 87501

Received on February 13, 2004; revised on May 26, 2004; accepted on May 28, 2004

Human and simian immunodeficiency viruses (HIV and SIV), influenza virus, and hepatitis C virus (HCV) have heavily glycosylated, highly variable surface proteins. Here we explore *N*-linked glycosylation site (sequon) variation at the population level in these viruses, using a new Web-based program developed to facilitate the sequon tracking and to define patterns (www.hiv.lanl.gov). This tool allowed rapid visualization of the two distinctive patterns of sequon variation found in HIV-1, HIV-2, and SIV CPZ. The first pattern (fixed) describes readily aligned sites that are either simply present or absent. These sites tend to be occupied by high-mannose glycans. The second pattern (shifting) refers to sites embedded in regions of extreme local length variation and is characterized by shifts in terms of the relative position and local density of sequons; these sites tend to be populated by complex carbohydrates. HIV, with its extreme variation in number and precise location of sequons, does not have a net increase in the number of sites over time at the population level. Primate lentiviral lineages have host species–dependent levels of sequon shifting, with HIV-1 in humans the most extreme. HCV E1 and E2 proteins, despite evolving extremely rapidly through point mutation, show limited sequon variation, although two shifting sites were identified. Human influenza A hemagglutinin H3 HA1 is accumulating sequons over time, but this trend is not evident in any other avian or human influenza A serotypes.

Key words: immune escape/*N*-linked glycosylation/neutralization antibody/variability/virus

Introduction

N-*linked glycosylation sites and their role in viral immune escape*

N-linked glycosylation sites are also referred to as sequons. For an asparagine (N) to be glycosylated, it requires the context of the amino acid pattern N-X-[S or T] (Marshall, 1974), where X can be any amino acid, followed by a Serine (S) or Threonine (T). A sequon will not be glycosylated if it contains or is followed by a Proline (Gavel and von Heijne, 1990), and glycosylation may be inhibited by certain combinations of N-X-S or when followed by specific amino acids (Kasturi *et al*., 1995; Mellquist *et al*., 1998; Shakin-Eshleman *et al*., 1996).

Alteration of a glycosylation site can have dramatic consequences for a virus. It can impact protein folding (Hebert *et al*., 1997; Land and Braakman, 2001; Slater-Handshy *et al*., 2004) and conformation (Meunier *et al*., 1999) and affect distant parts of a protein through masking or conformational changes. Although gain of a carbohydrate can sterically mask epitopes, the loss of one could result in tighter packing of glycoprotein regions involved in neutralization epitopes, reduce accessibility, and so also facilitate immune escape (Ye *et al*., 2000). The loss of sequons can even impact immunogenicity of noncovalently associated proteins, for example a change in sequons in the human immunodeficiency virus type 1 (HIV-1) transmembrane envelope (Env) protein gp41 induces conformational changes in the associated Env gp120 surface protein that dramatically diminish the binding of many gp120- specific antibodies (Si *et al*., 2001). Altered patterns of glycosylation in viral proteins can also contribute to escape from T cell responses (Botarelli *et al*., 1991; Ferris *et al*., 1999; Selby *et al*., 1999) and influence receptor binding and phenotypic properties of viruses (Kaverin *et al*., 2002; Koito *et al*., 1995; Matrosovich *et al*., 1999; Ogert *et al*., 2001; Pollakis *et al*., 2001).

Influenza, glycosylation, and antigenic drift

Some of the earliest studies on the biological and immunological consequences of glycosylation site variation were conducted in influenza proteins (Alexander and Elder, 1984). The number of sequons in the heavily glycosylated influenza A hemagglutinin 1 (HA1) of the pandemic H3 virus has increased from 6 to 10 since it entered the human population in 1968 (Skehel and Wiley, 2000), and the increase is assumed to make HA1 generally more refractive to antibodies. For example, the amino acids around the *N*-linked glycosylations site at position N165 of HA stopped participating in antigenic drift (Skehel and Wiley, 2000; Wiley and Skehel, 1987).

HIV Env and glycosylation site variation

HIV-1 is highly variable, and variants are grouped through phylogenetic analysis into major clades or subtypes (A–K). Recombinant forms of HIV are very common (Robertson *et al*., 2000), and when a lineage based on recombination

[1]To whom correspondence should be addressed; e-mail: btk@lanl.gov

between two subtypes becomes an important epidemic lineage, it is called a circulating recombinant form (CRF). HIV-1 varies dramatically within a clade and within infected individuals. HIV Env gp120 is among the most heavily glycosylated proteins in nature (Myers *et al.*, 1992), far more heavily glycosylated than envelopes of other retroviruses of similar size (e.g., HTLV-1, MuLV) (Polonoff *et al.*, 1982). The influence of sequon variation on HIV antibody recognition and viral phenotype has been studied in the context of HIV and SIV Env proteins (Chackerian *et al.*, 1997; Cheng-Mayer *et al.*, 1999; Losman *et al.*, 2001; Ly and Stamatatos, 2000; Malenbaum *et al.*, 2000; Matthews *et al.*, 1987; Ratner 1992; Ye *et al.*, 2000). Sigvard Olofsson and colleagues first showed that glycosylation of gp120 could change exposure of neutralizing antibody epitopes (Bolmstedt *et al.*, 1996). In the HIV-1 CRF01, the absence of a sequon near the base of the V3 loop, an important antigenic doman of Env gp120, was correlated with rapid amino acid substitutions and positive selection (Kalish *et al.*, 2002), suggesting a similar situation to the influenza A HA protein, where a glycosylation site may provide regional protection from antibodies (Skehel and Wiley 2000). The number of Env sequons in both HIV and SIV infections varies extensively within infected individuals (Overbaugh and Rudensey, 1992; Wolinsky *et al.*, 1992), and this variation constitutes a mechanism of immune escape during the course of an HIV or SIV infection (Cheng-Mayer *et al.*, 1999; Davis *et al.*, 1990; Simmonds *et al.*, 1991).

A heavily glycosylated face of the 3D structure of Env gp120 has been called the immunologically silent face (Moore and Sodroski, 1996). Generally carbohydrate moieties appear as self to the immune system, so this face reduces the antigenicity of a large region on the gp120 surface. Glycosylation of variable loops restricts access to conserved receptor binding sites and limits their exposure to the immune system (Wyatt and Sodroski, 1998), and HIV Env has been described as having a glycan shield (Wei *et al.*, 2003).

A recent survey of the global collection of HIV-1 Env gp120 surface protein M group sequences in the Los Alamos HIV database showed that gp120 varied in length between 484 and 543 amino acids (Korber *et al.*, 2001). The number of potential sequons in gp120 ranges between 18 and 33 with a median of 25 (Korber *et al.*, 2001) (Figure 1) (this range is often ignored and HIV is frequently reported to have 25 *N*-linked glycosylation sites). The dramatic variation in number of sequons partly results from gp120 length variation frequently involving insertions and deletions of potential glycosylation sites inside HIV hypervariable domains, however, evolutionary propensities in base substitution patterns in HIV may also contribute to the rapid flux in numbers of sequons in gp120 (Bosch *et al.*, 1994; Kuiken *et al.*, 1999).

Not all potential glycosylation sites on a given HIV-1 Env protein are fully occupied (Zhu *et al.*, 2000). For example, NNTT is a common pattern among HIV sequences, and steric occlusion may prevent carbohydrate addition to both asparagines, and the protein context of sequons does not always favor glycosylation (Kasturi *et al.*, 1995; Mellquist *et al.*, 1998; Shakin-Eshleman *et al.*, 1996).

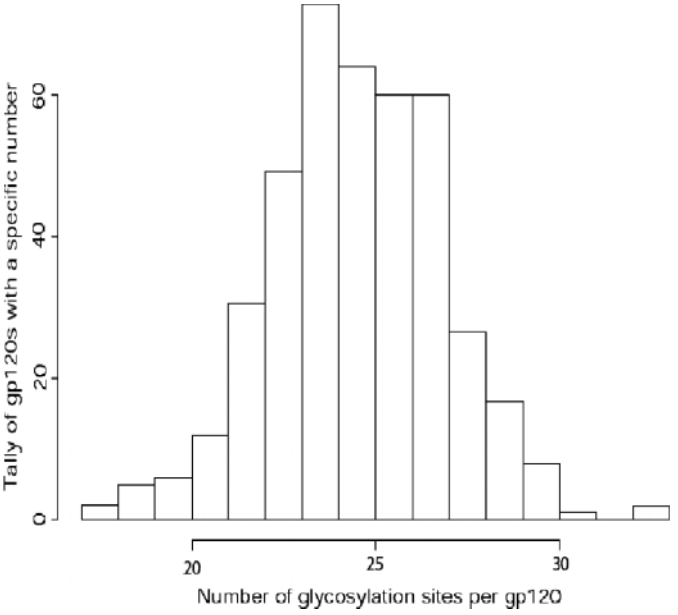

Fig. 1. Histogram showing the relative frequency of envelope gp120 proteins with different numbers of N-linked glycosylation sites found in the HIV database (www.hiv.lanl.gov, 2002 listing). Only one sequence obtained from a given individual and complete gp120 Env sequences were included in this comparison ($n = 386$).

Glycosylation and hepatitis C Env diversity

Hepatitis C virus (HCV) belongs to the Hepacivirus genus in the Flaviviridae family (Rice, 1996). Like HIV-1, it is highly variable; HCV establishes a chronic infection in most hosts and so is subject to continuous immune pressure and rapid accumulation of mutations. The virus has been classified into six different genotypes, which have each been subdivided into a large number of subtypes. The Env E1 and E2 proteins of HCV form heterodimers on the virion surface, and glycosylation is essential for this dimerization (Meunier *et al.*, 1999). The variability of both proteins is comparable, from 88% nucleotide (90% amino acid) identity between strains from the same subtype to 55% nucleotide (59% amino acid) identity between different genotypes. In both E1 and E2, *N*-glycosylation sites are limited to the amino terminus of the proteins; the carboxy-terminal region of these proteins is the transmembrane portion. The efficient glycosylation of E1 is dependent on the presence of E2 in a polyprotein (Deleersnyder *et al.*, 1997), although it does not appear to depend on the specific sequence of E2 (Dubuisson *et al.*, 2000), and the noncovalent association of E1 and E2 depends on the first and fourth glycosylation sites of E1 (Meunier *et al.*, 1999). It has been shown that the glycans attached to the E1E2 heterodimer prior to budding of the virus are exclusively high-mannose (Deleersnyder *et al.*, 1997), although E1E2 found circulating on HCV virions also have complex carbohydrates (Sato *et al.*, 1993). As in HIV, sequon changes have been shown to change antibody exposure in HCV (Fournillier *et al.*, 2001).

In this study, we characterize patterns in viral glycosylation site variation at the population level for influenza, HIV, and HCV and describe a Web-based program that

was used to facilitate tracking sequons in protein alignments for these comparisons.

Results

Influenza proteins and N*-linked glycosylation patterns over time*

Using alignments and sampling dates obtained from the 2003 Los Alamos Influenza Sequence Database (www.flu.lanl.gov), we confirm (Figure 2A) the H3 HA1 protein gradually increases in the median number of *N*-linked glycosylation sites over time (from 6 to 10), with a small amount of within-year variation. The acquisition of some of these sites can be directly related to antigenic drift (Skehel and Wiley, 2000), the change in the antigenic profile of influenza from year to year that necessitates annual review and frequent updates of the vaccine strain.

However this net increase in the number of sequons is not a general feature of influenza evolution and is only found in the human H3 serotype of influenza A, not in other serotypes of influenza A (Figure 2), nor in the avian H3 over time, or in any other avian serotype with adequate sampling over time, nor in the human influenza B HA1 (Figure 2, plus summary in legend). There was also no trend for increasing numbers of sequons over time in the avian or human neuraminidase (N2) serotype proteins, although this protein is also heavily glycosylated and shows variable numbers of sites from year to year (Figure 2D).

Global trends in glycosylation patterns in HIV Env

There is no net tendency for the sequon number to increase or decrease over time in the HIV-1 M group, or within subtypes or CRFs (Figure 3). Two kinds of sequons are evident in HIV-1 and HIV-2: those embedded in readily aligned positions and those embedded in hypervariable regions that shift in relative position and regional density by base mutation and by insertions and deletions. Here we refer to these two classes of sequons as *fixed* and *shifting*, respectively. Although most of the shifting sequons are found in variable loops, some are also present in the C4 region (conserved, or C, domains in the HIV envelope were called conserved because they are relatively conserved when compared to the variable regions, however, they can also span insertions and deletions that can result in shifting sequon locations). The location and frequency of sequons in gp120 in each major subtype of the HIV-1 M group (Gao *et al*., 1996; Robertson *et al*., 2000) are illustrated in Figure 4. Despite the extreme variability between isolates, there is an essentially conserved pattern of variation in each of the HIV-1 M group subtypes, and even in the genetically very distant HIV-1 group O. (There are insufficient full-length HIV-1 group N sequences for a comparison). Protein regions show the same frequencies for most sequons in each HIV-1 lineage and subtype, suggesting that selective pressures on sequons in these diverse lineages are consistent (Figure 4).

HXB2 and SF2 are common HIV-1 reference strains for which the *N*-linked carbohydrates additions to Env gp120 have been biochemically defined (Leonard *et al*., 1990; Zhu *et al*., 2000). We aligned the SF2 protein sequence to HXB2 strain and noted the types of carbohydrate found: mannose, complex carbohydrate, or a mixture of both (hybrid carbohydrates). We also found that different types of carbohydrates were associated with fixed and shifting sequons defined at the population level (Figure 4). There are 24 *N*-linked glycosylation sites in HXB2, and 26 in SF2, and they differ in location in seven cases. In HXB2, all seven shifting sequons in Env gp120 are complex carbohydrates (7/13 complex carbohydrate additions, 54% are in the shifting regions), whereas all high-mannose or hybrid carbohydrates are in fixed positions (11/11, 100%, Fisher's exact test p-value = 0.006). When an additional sequon in the shifting V1 region of the HXB2-related variant LAI was later characterized, it also carried a complex carbohydrate (Gram *et al*., 2002), consistent with this pattern. For SF2, some of the sequon positions contained hybrid carbohydrates, and some were only partially glycosylated. Though the high-mannose residues were not significantly correlated the fixed sequons of SF2, the trend from HXB2 was evident: 8/11 (72%) of high mannose glycosylation sites were in fixed sequons, whereas 8/14 (57%) complex carbohydrates were in shifting sequons, and one was unknown. A structural model of the fully glycosylated SF2 gp120 suggests that the high-mannose glycans are clustered on one surface of the protein and the complex carbohydrates are localized on distinct region of the protein surface (Zhu *et al*., 2000).

The 2G12 monoclonal antibody is one of the few broadly cross-reactive human HIV-neutralizing antibodies, and it has an unusual epitope involving several mannose carbohydrates that project from the generally immunorefractive glycosylated face of HIV-1 Env. Not surprisingly, cross-competition studies have shown that B cell responses to 2G12 epitope are very unusual in HIV-1-infected individuals (Moore and Sodroski, 1996). The sites that comprise the epitope, indicated in Figure 4, require mannose at positions of N295, N332, and N392 for 2G12 binding (Sanders *et al*., 2002; Scanlan *et al*., 2002). (See the HXB2 sequence locator tool in the HIV database [www.hiv.lanl.gov] to determine the specific positions referred to in this text.) These sites are well conserved in most subtypes, and tend to show comparable levels of glycosylation. The exception is subtype C, which only rarely has a glycosylation site immediately next to the Cys at the base of the V3 loop (N295). It has been suggested that the 2G12 epitope may be conserved because its structure enhances gp120–mannose interactions with the human protein dendritic cell–specific HIV-1-binding protein (DC-SIGN), an interaction that faciliatates efficient HIV-1 infection (Sanders *et al*., 2002).

The global collection of HIV Env gp120 molecules shows small but significant variations in the distributions of sequon frequencies for each subtype (Kruskal-Wallis p-value = 0.0002). Occasionally a site will be completely lost or added in a subtype (Figure 4), like the loss of a shifting sequon in CRF01 (subtype E in Env) in the V4 region or the additional *N*-terminal fixed site in O group. These differences may simply reflect a founder effect in the lineage, or may confer a critical change to the conformation of Env in the context of a particular lineage. Although the sequon frequencies in different HIV-1 subtypes are

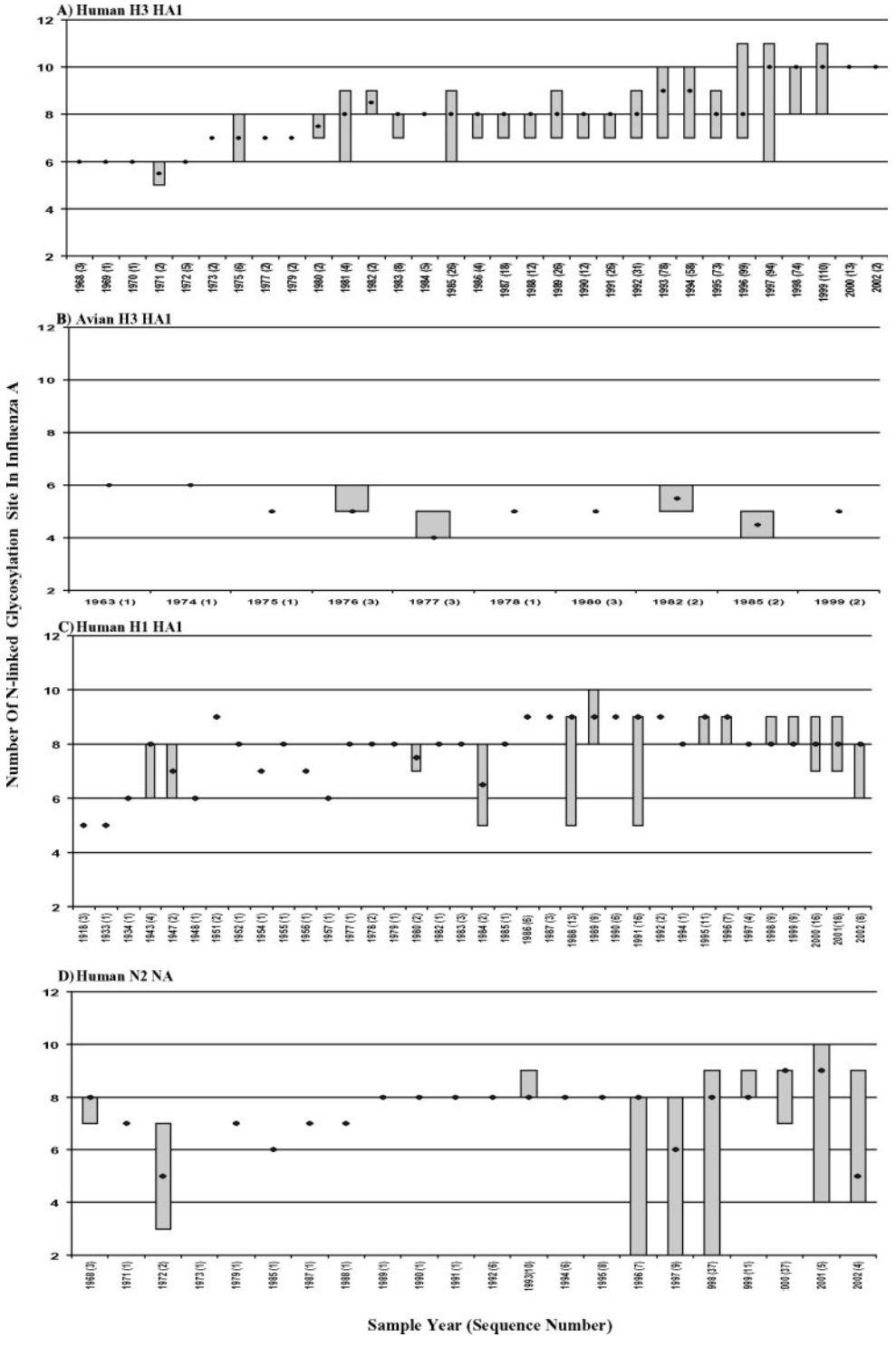

Fig. 2. Changes in the number of N-linked glycosylation sites in different influenza A proteins over time. (**A**) Human H3 HA1 proteins. This figure shows the median value and the range of number of sequons per sequence by year of sampling, with the number of sequences from each year in parentheses and the year noted on the *x*-axis. Samples were collected between 1968 and 2002. The increase in number of sites each year is evident; the *p*-value indicating this accumulation is not due to chance alone is very low ($p \ll 10^{-5}$), however, this is misleading because the epidemic strains in any one year are not independent (Korber *et al.*, 2001). (**B**) Avian H3 HA1 proteins. The H3 HA1 avian samples collected between 1963 and 1999 oscillate between four and six glycosylation sites. (**C**) Human H1 HA1 proteins. The H1 HA1 human isolates obtained between 1934 and 2002 oscillate between five and nine sites, with a lone samples from 1918 and 1933 having only five sites. (**D**) Human neuraminidase N2 proteins. The N2 proteins from human isolates do not show accumulation of glycosylation sites over time, although there is variation (between two and nine sites). Other influenza HA1 molecules with adequate data for testing are not shown because no clear trends over time were apparent. These include the avian H5 HA1, oscillating between five and eight sites between 1959 and 2001 ($n = 94$); avian H7 HA1, oscillating between three and five sites between 1927 and 2000 ($n = 122$); avian H9 HA1, staying steady with a median of six sites between 1966 and 2001 ($n = 63$), and the human influenza B hemagglutinin HA1, which maintained a stable median of seven sites with a range of six to eight from 1940 to 2002 ($n = 420$).

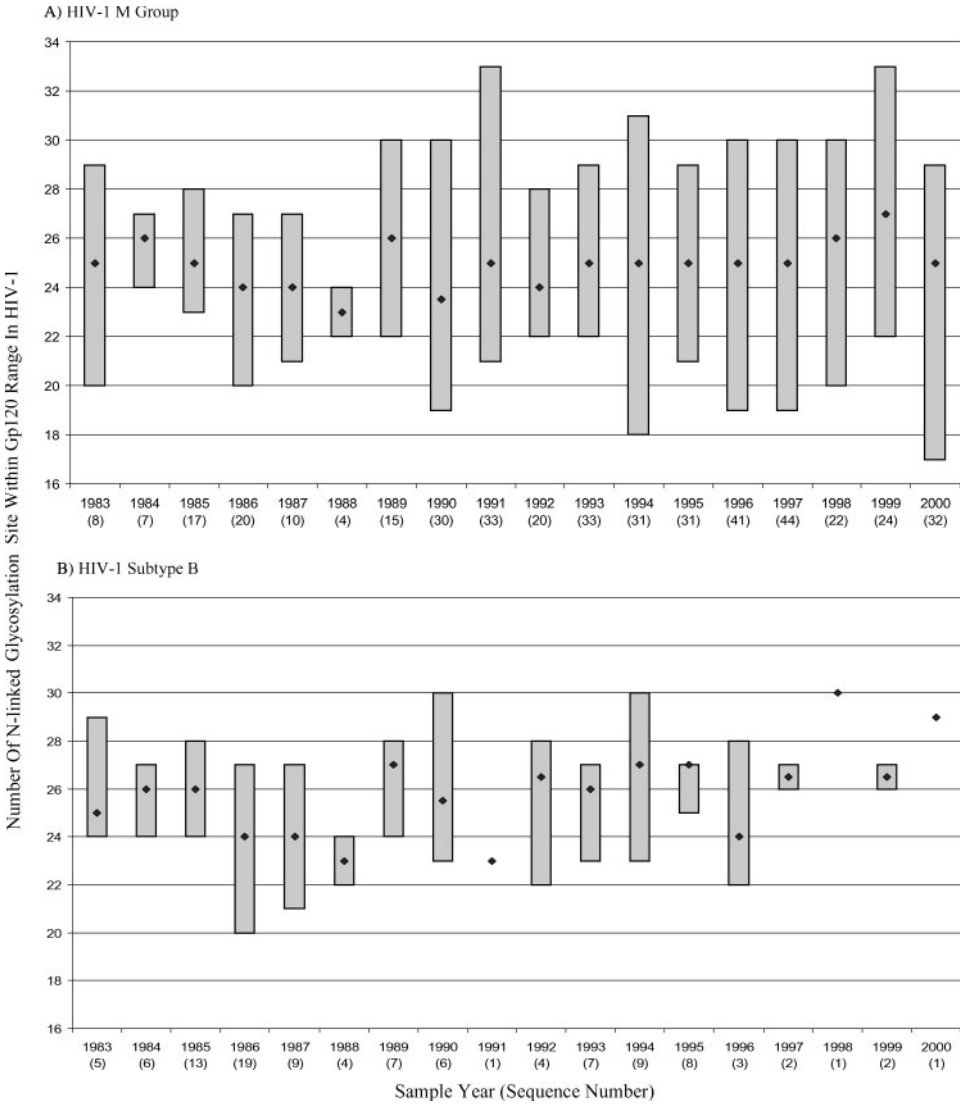

Fig. 3. The median and range of number of *N*-linked glycosylation sites per Env gp120 in the HIV-1 M group and subtype B between 1983 and 2000. (**A**) No correlation was observed in terms of increasing or decreasing numbers of sequons in gp120 at the population level over a 20-year period of sampling of M group sequences in the database. (**B**) Distributions of the number of *N*-linked glycosylation sites for each of the subtypes and CRFs in **A** did not reveal any trends in terms of accumulation or decline in number of sequons over time; subtype B is shown as a representative set.

significantly different, perhaps more important are the similarities between subtypes: The range of sequon frequencies is broad and basically overlapping within all subtypes (Gao *et al*., 1996) (Figure 5), and the conserved sites as well as frequencies of variable sites tend to be comparable between subtypes (Figure 4).

N-*linked glycosylation patterns related to coreceptor usage and phenotype*

HIV-1 Env enters CD4-positive T cells through a series of steps involving binding to both the CD4 protein and a chemokine receptor protein, usually CCR5 (R5 viruses) or CXCR4 (X4 viruses); some viruses can use either chemokine receptor (R5X4 viruses) (Cho *et al*., 1998). Sets of Env protein sequences derived from R5, X4, or R5X4 viruses were obtained from in the Los Alamos HIV database (www.hiv.lanl.gov). The total number of sequons in gp120, and in subregions of gp120 including the variable domains V1 and V2, which play a role in coreceptor usage in certain contexts (Cho *et al*., 1998; Hoffman *et al*., 1998; Pollakis *et al*., 2001), were nearly identical among R5, X4, and R5X4 viruses. However, two sequons in fixed sites show distinctive frequencies (Figure 6). The most striking was the sequon located within the V3 loop at position N300 of HXB2, which was present in all 44 R5 isolates (100%), but only in 4/11 (36%) X4 isolates and 7/11 (64%) R5X4 isolates (Fisher's exact test p-value $= 4 \times 10^{-7}$). Site-directed mutagenesis has shown this site to be associated with coreceptor usage (Ogert *et al*., 2001; Pollakis *et al*., 2001). The other potentially interesting site was at position N230 of HXB2, which was less common in the CCR5-utilizing group relative to the CXCR4 and R5X4 viruses, although this difference was just a trend and not statistically significant (11/44 [25%] had the site among R5 viruses, versus 6/11 [55%] in X4, and 5/11 [45%] in R5X4).

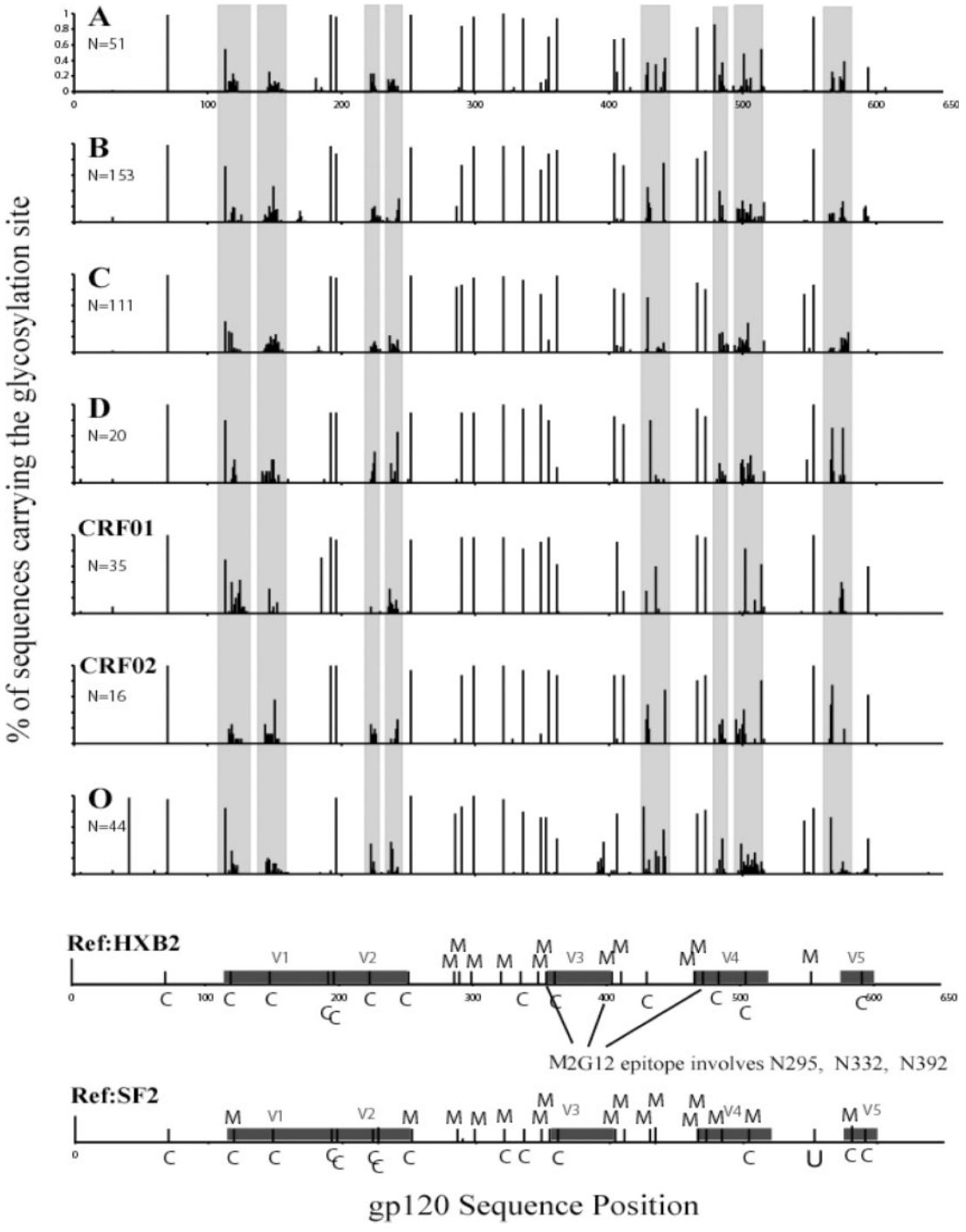

Fig. 4. Frequency of *N*-linked glycosylation sites in HIV Env gp120 based on an alignment of HIV-1 M group subtypes and circulating recombinant forms. Sequons in HIV-1 subtypes and CRFs with the most available full-length gp120 sequences are shown here. The sequences were aligned, and the fraction of sequons in every position in the alignment of each subtype is indicated. The sequons we refer to as shifting are highlighted in gray. The alignment is numbered according to positions in the HXB2 reference strain. The sequons in particularly positions in two HIV-1 reference strains are indicated, HXB2 and SF2. The biochemistry of the oligosaccharide additions has been determined for these two strains (Leonard *et al*., 1990; Zhu *et al*., 2000), and high mannose is indicated at the top of appropriate sequons with an M, complex carbohydrates are labeled under each sequon with a C. Some mixtures of both were identified in SF2 (Zhu *et al*., 2000). The locations of variable loops in gp120 are indicated in black box on the reference strains. The high-mannose carbohydrates critical for 2G12 binding are indicated (Sanders *et al*., 2002; Scanlan *et al*., 2002).

A virus's ability to form multinucleated syncytia in indicator cell lines distinguishes R5 and X4 viruses. Viruses isolated from newly infected people tend to be nonsyncytium-inducing (NSI) and R5, whereas X4 and syncytium-inducing (SI) viruses tend to appear later in infection, and the SI/NSI designations are available for many viral sequences that do not have defined coreceptor usage. So to further compare the glycosylation patterns

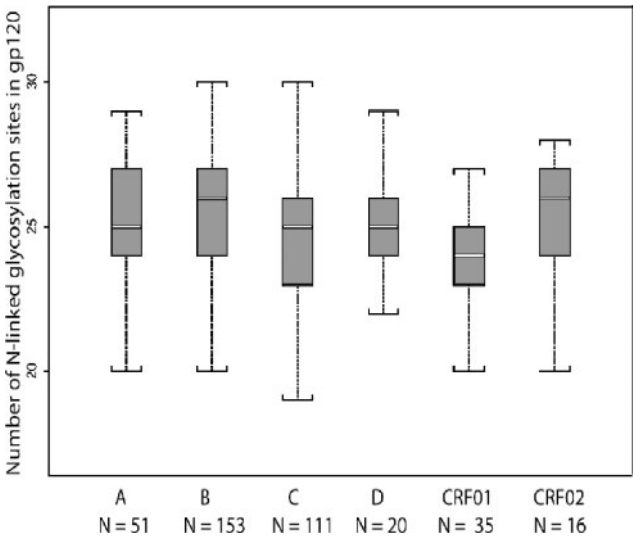

Fig. 5. Distributions of number of *N*-linked glycosylation sites in HIV-1 Env gp120 proteins by HIV-1 subtype. A box plot indicating median and interquartile range of the number of sequons in the most commonly sequenced subtypes and circulating recombinant forms of HIV-1 M group is shown. The number of full-length sequences included in the summary of each clade is noted. A nonparametric Kruskal-Wallace test indicated that there were distinction between the distribution of numbers of sequons in the M group subtypes and circulating recombinant forms (*p*-value of 0.0002), despite the overlap in the distributions.

with viral phenotypes, we analyzed the number and position of sequons in full length Env, including gp41 (Figure 7) as well as gp120, from SI and NSI patients. Two significant changes were observed: One was the loss of a sequon at position N136 in SI viruses, located in the V1 loop, and the other was a gain of a sequon in gp41 at position N674.

We then determined if the sequon alterations associated with phenotypic change at the population level are mirrored in an individual patient (Hu *et al*., 2000). In the patient studied, there was a one-to-one correspondence between NSI and R5 usage and SI and X4 usage. All significant changes of sequons were found in regions V1, V2, V4, and V5, primarily in the *N*-terminal regions of V1 and V4. In gp120, there are 19 highly conserved sites, 4 additional sequons from NSI to SI, and 5 losses from NSI to SI. Three of these changes were in shifting sites, simply moving the sequon by one position relative to context of the protein (in V1, N136 to N136+, N141 to N142, and in V4, N401 to N402). The NSI-to-SI phenotypic switch was accompanied by these three shifts in sequon position, a loss of sequon N188 in V2 and N405 in V4, and a gain of a sequon at position N463 in V5. In this patient, the sequon at position N300 was unchanged and not related to coreceptor usage, and gp41 sequons were invariant, emphasizing that there are exceptions to the statistical correlations that can be detected at the population level.

N-*linked glycosylation patterns in other primate lentiviruses*

Among the primate lentiviruses, Env sequon variation in M group viruses is the most extreme relative to the genetic distances based on base substitution (Wills *et al*., 1996). However, SIVs do vary extensively and also show some interesting glycosylation patterns (Chackerian *et al*., 1997; Nyambi *et al*., 1997; Overbaugh and Rudensey, 1992). Using phylogenetic analysis, we selected HIV-2 and SIV sequences representing the spectrum of natural diversity in each lineage. Shifting sites, analogous to those found in HIV-1, were apparent in HIV-2 and in chimpanzee viral Envs (SIV_{CPZ}). The human HIV-1 epidemic is thought to have resulted from a cross-species transmission from chimpanzee, whereas human HIV-2 and macaque SIV_{MAC} result from cross-species transmission of SIV_{SMM} virus from the sooty mangabey, its natural host (Gao *et al*., 1996; Hirsch *et al*., 1989). In SIV_{SMM} and African green monkey SIV_{AGM}, shifting sites tended to resolve into fixed sites (Figure 8A), with the exception of the V4 region in SIV_{AGM}. Table I highlights the sequons in the V1 loops from sequences used to generate the frequencies shown in Figure 8. Although there was length variation in SIV_{SMM} and SIV_{AGM} V1 sequences, the relative placement of the sequons and the Cys involved in the disulfide bridge that forms the base of the loop was conserved (Table I). In contrast, shifting sequons are frequently seen in infected macaques (SIV_{MAC}) followed over time through progression to simian AIDS (Chackerian *et al*., 1994, 1997; Overbaugh *et al*., 1991) and in HIV-2 sequences in human (Figure 8A, Table I). Because HIV-2 and SIV_{MAC} infections result from cross-species transmission of SIV_{SMM} virus from sooty mangabeys (Gao *et al*., 1996; Hirsch *et al*., 1989), the degree of variation in shifting sites is host-specific for viruses of this lineage. SIV_{SMM} does not cause disease in sooty mangabeys, but it does in macaques, as does HIV-2 in humans. Like SIV_{SMM}, SIV_{AGM} does not cause disease in its natural host species, African green monkeys, and is readily found in animals in the wild (Norley *et al*., 1999; Ohta *et al*., 1988).

HCV E1 and E2 protein N-*linked glycosylation site patterns*

The sequons in HCV Env E1 and E2 proteins are far less variable in HIV-1 envelopes, despite HCV otherwise being an extraordinarily variable virus. After creating alignments of the available HCV E1 and E2 sequences in GenBank, the sequences were checked by phylogenetic analysis, and very similar sequences were removed, leaving 294 distinct E1 protein sequences and 130 E2 sequences for comparisons. Numbering of the E1 and E2 positions in this section is based on the Los Alamos Hepatitis C Database HCV Sequence Locator tool (http://hcv.lanl.gov/content/hcv-db/LOCATE/locate.html).

There are typically 5–7 sequons in E1 over a stretch of 194 amino acids. Four of these sites are highly conserved among all genotypes. The site at position N61 in the alignment was present in genotype 6 and 1b but absent in all other genotypes (Figure 9, top). A pair of adjacent sequons at positions N43/N44 also showed subtype-specific variation, the site at position N43 being present in subtype 2b, and the one at N44 in 2a and 2c, as well as in most other sequences. In 35% of the subtype 1b sequences, the N43 and N44 sites were both present; this occurred in 1 sequence of another genotype, a 3b. The sequences with two sequons did not

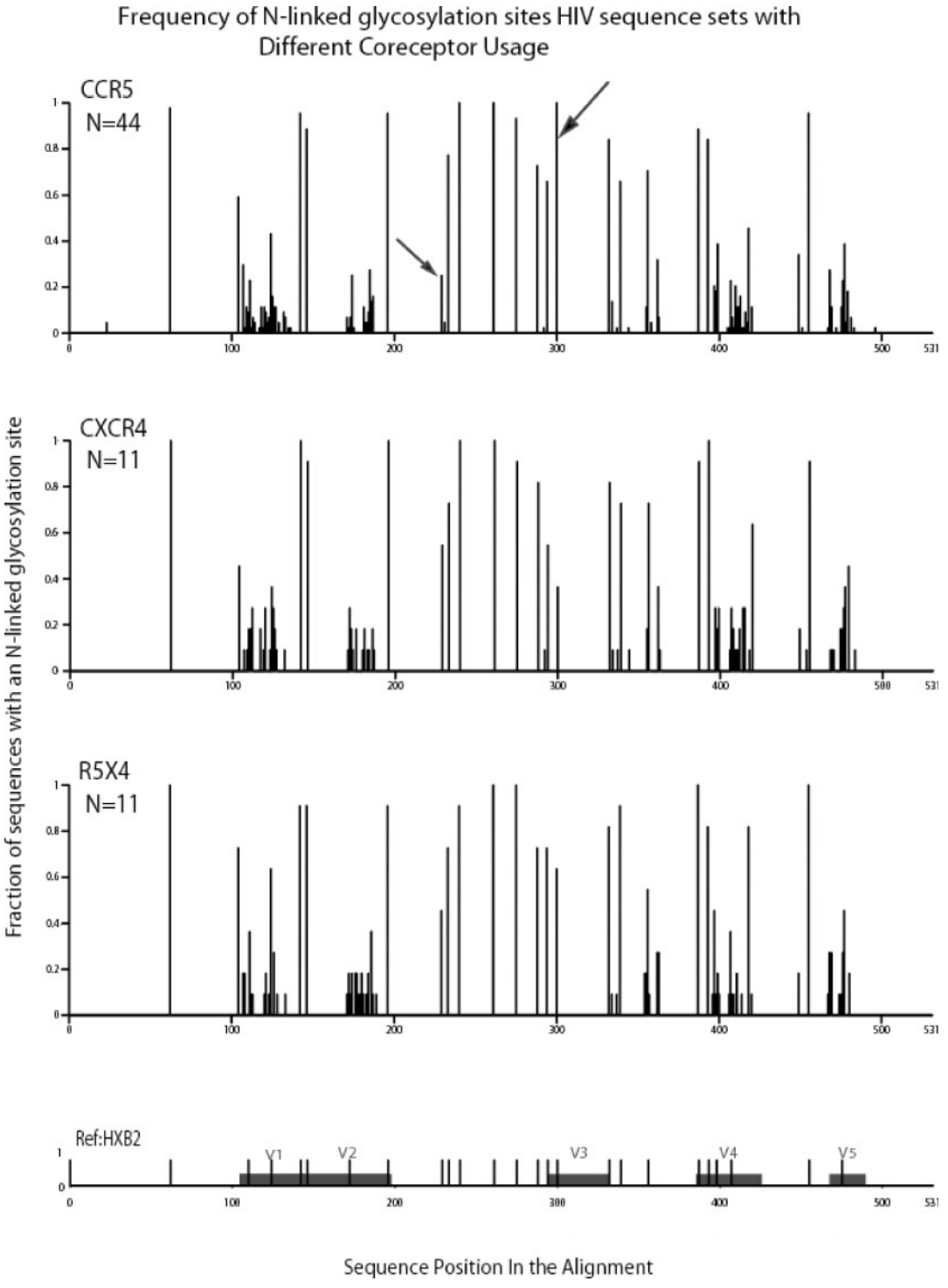

Fig. 6. The relationship between coreceptor usage and *N*-linked glycosylation sites in the HIV-1 M group. The sequences that made up the sets associated with particular coreceptor usage were very diverse M group sequences from a variety of subtypes. The set of 44 R5 viruses included 8 subtype A, 12 B, 13 C, and 11 others; the 11 X4 viruses included 1 subtype A, 5 B, 2 C, and 3 D; the 11 R5X4 viruses included 3 subtype A, 6 B, 1 C, and 1 D. Arrows in the R5 graph indicate the sequons with potentially distinctive frequencies.

appear to come from a specific geographic region and probably arose independently. The sequons at positions N5 and N136 of E1 have been shown to be essential for formation of E1–E2 complexes on the virion surface (Meunier *et al.*, 1999); N136 also is important in reducing the antigenicity of the virus (Fournillier *et al.*, 2001). Little is known about genotype- or subtype-specific differences in viral phenotype.

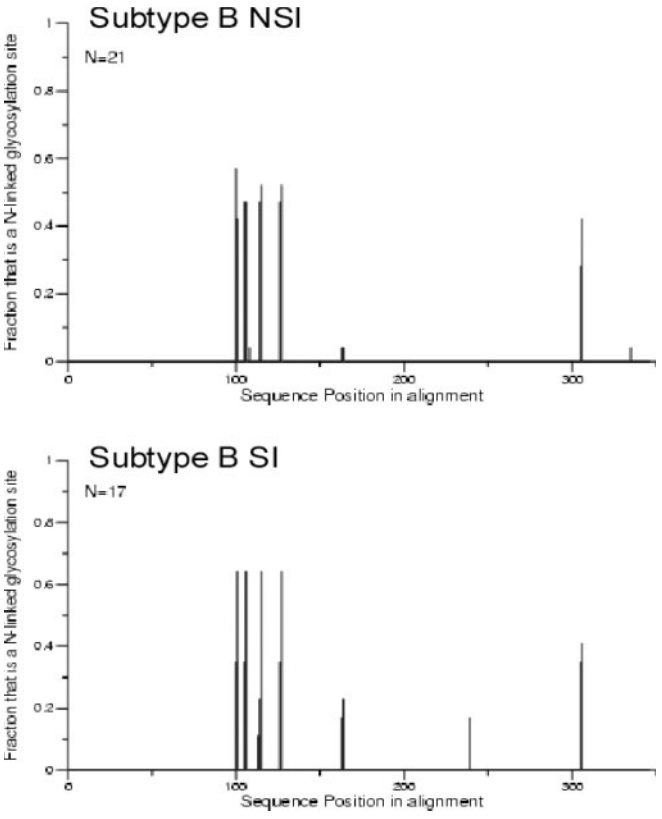

Fig. 7. *N*-linked glycosylation site changes in Env gp41 in NSI isolates and SI isolates from patients infected with HIV-1 subtype B. HIV-1 B gp41 sequences were aligned and analyzed for sequon differences between 21 NSI and 17 SI sequences. Protein sequences were obtained from the HIV database; only one sequence per patient was used. Sequences were analyzed for patterns of glycosylation using the LANL *N*-glycosite program (http://hiv-web.lanl.gov/content/hiv-db/GLYCOSITE/glycosite.html).

Sequons in the E2 protein were mostly limited to the *N*-terminal part of the protein, with 10–11 sites all located before amino acid 272 (the total length of the E2 alignment was 436 amino acids). Some type-specific sequon patterns were also found in E2, as all sequences of genotypes 3 and 6 missed the site at position N160. Only one or two sequences of genotypes 4 and 5 were available, so variation analysis was not possible for these genotypes. Within genotype 1, the sequon at position N94 was highly conserved in subtype 1a but present in only 17/101 sequences of subtype 1b. This site was one of the two sequons in E2 that shifted in terms of its relative position, that is, it could be found starting anywhere between positions 92 and 95 (Figure 9, bottom), suggesting that either there is pressure for this subtype to incorporate a glycosylation site in this region of the protein or that there is a selective advantage in retaining the site in the region but modifying its precise location. Another shifting site was found in E2, specifically in genotype 6. This site was embedded in the only region with multiple insertions and deletions in HCV, and it shifts relative location on the basis of insertions and deletions, and is lost in one of the sequences (Figure 9, bottom). The presence of shifting glycosylation sites in the E2 protein, like HIV-1 Env, is probably related to its profound immunogenicity and escape-related variability.

Discussion

The potential of changing *N*-linked glycosylation sites as a mechanism for immune evasion has long been appreciated (Alexander and Elder, 1984). In influenza, the steady increase in the number of sequons in H3 HA1 over time (Skehel and Wiley, 2000) is the exception, as the number of sequons in human H1 HA1 and neuraminidase N2 proteins oscillates but does not show a net increase. The sequon frequency in avian influenza hemagglutinin proteins oscillates at roughly comparable levels to the human proteins, despite having a very different ecology: In aquatic avian species flu is asymptomatic, and the viruses are in evolutionary stasis (Webby and Webster, 2001), whereas in human epidemics, the virus causes disease and generates new variants that are able to infect hosts who have become immune to earlier influenza strains (Bush *et al.*, 1999; Ferguson *et al.*, 2003). It seems possible that the increase in sequon frequency in H3 HA1 in humans is a local fluctuation over time, and it too will move to random oscillation in future years. Although year of sampling was not readily available for the HCV sequences, HCV shows much less variation in sequons, and this variation seems to be due to lineage-specific patterns that would be unlikely to vary by sampling year.

There is a striking recapitulation of sequon patterns in different levels of HIV evolution. The level of sequon variation that can be seen in an individual is mirrored at the level of variation seen within a subtype, which is further reflected at the level of variation seen within the HIV-1 M group global epidemic. Thus the boundaries of the potential to add and eliminate sequons to facilitate immune escape and infect a range of cell types, within what is tolerated in terms of fitness costs, may be explored anew during the course of each HIV-1 infection of a single individual. Although the phylogenetic tree of HIV-1 is clearly expanding over the time period during which we have followed HIV-1 in the human population (Korber *et al.*, 2000; Robbins *et al.*, 2003; Yusim *et al.*, 2001), the sequon variations in gp120 and gp41 seem instead to be rapidly fluctuating within inherent boundaries.

There is probably an upper bound to the number of sequons that can be maintained on Env gp120, because carbohydrate structures are quite large and their presence or absence must influence the protein's conformation. For example, there are two sequons next to cysteines that form the base of the V3 loop. A typical glycan is about 2000 Da, and the V3 loop is only about 3000 Da, so the presence or absence of these sequons would logically impact the orientation with which the V3 loop projects from the protein surface. Glycosylation sites may be preserved *in vivo* to mask neutralizing antibody sites, but replication efficiency and access to receptor binding sites may provide a counterbalancing force—multiple glycosylation sites can be removed from gp120 without loss of infectivity (Ohgimoto *et al.*, 1998).

The observation that the complex carbohydrates tend to be localized in the shifting positions while high mannose is

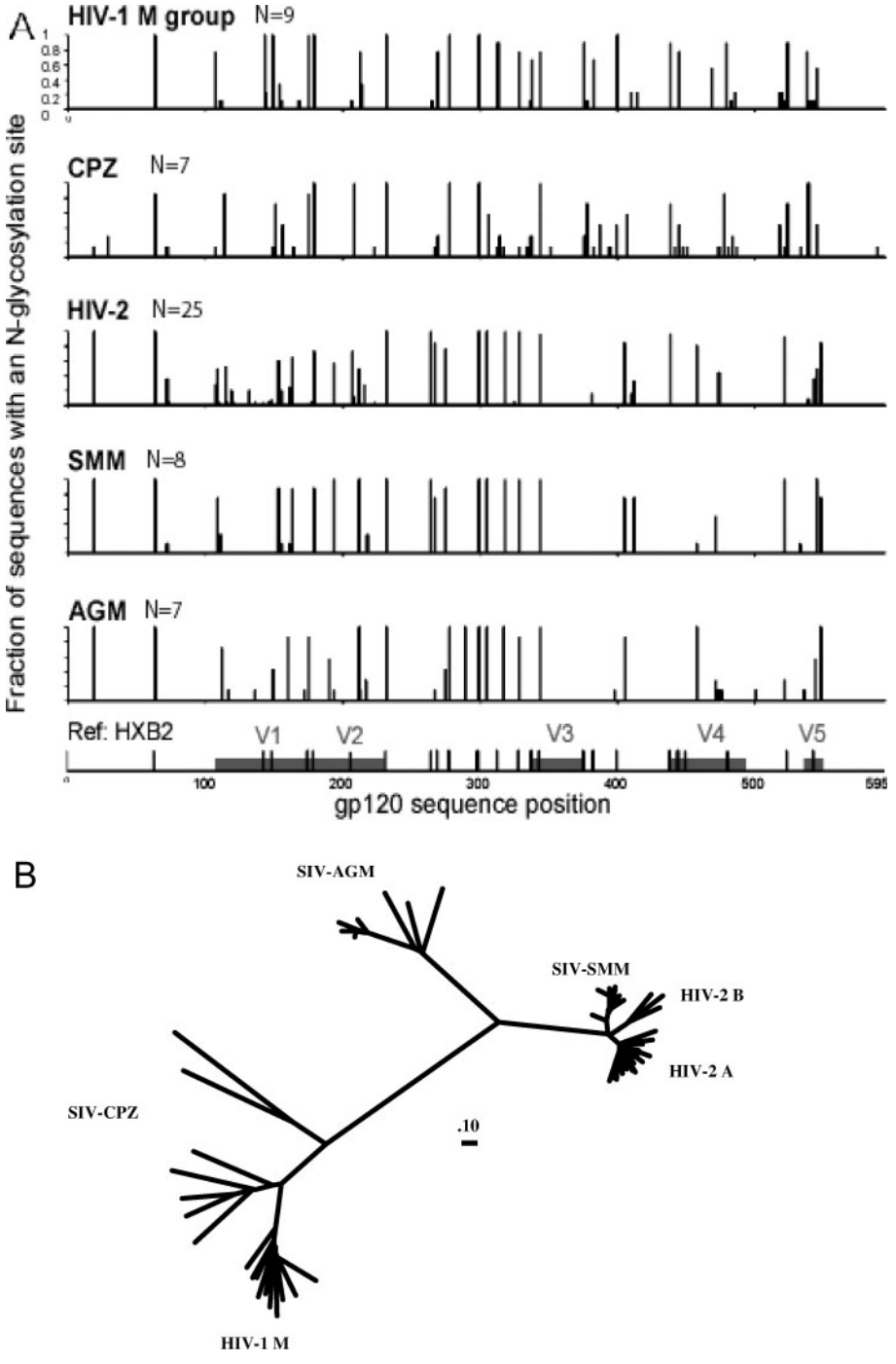

Fig. 8. (**A**) Sequon frequencies in different primate lentiviral lineages. Both shifting and fixed glycosylation site patterns are readily identified in the HIV-1 M group, CPZ, and HIV-2 sequences. Although only limited numbers of complete gp120s that represent distinct lineages of SIV_{SMM} and SIV_{AGM} are available, the shifting sites are not readily apparent in these viruses when infecting their natural hosts. Two exceptions are that shifting sequons are not evident V4 in the HIV-2, and are indicated in the SIV_{AGM} set. (**B**) Phylip PROTDIST tree of gp120 sequences (http://evolution.genetics.washington.edu/phylip.html) used for the analysis shown in **A**. This tree is simply meant to illustrate the extent of diversity in the population of protein sequences of these viruses in different hosts. Most of the specific sequences used to generate **A** and **B** of this figure are listed in Figure 4, although only a subset of the HIV-2 sequences are included in Table I to save space (18 are included here).

favored in the fixed positions in HIV-1 Env gp120 could result from multiple contributing factors. The interactions of DC-SIGN and other C-type lectins with gp120 high-mannose oligosaccharides have an important role in HIV infectivity (Lin *et al.*, 2003), suggesting there may be a fitness cost in disrupting their precise orientation. If the high-mannose oligosaccharides are important for successful sexual transmission, there would be selection pressure to

Table I. Relative spacing of unaligned V1 sequons to illustrate the level of diversity found in shifting sites

Sequence name	No. *N*-linked glycosylation sites in V1	No. sites in full gp120	Spacing of sequons in V1 loops
A) HIV-1: One patient over time			
1988	NA	NA	C....N..N.............NC
1994	NA	NA	C.....N..N.............NC
1994	NA	NA	C.....N..N..N..N..N.................NC
1994	NA	NA	C....N..N.............NC
1995	NA	NA	C....N.....................NC
1995	NA	NA	C....N.......N..N...........NC
1995	NA	NA	C.....N......N..........NC
B) HIV-1: A subtype			
HIV1_A_MA246	4	25	NC.....N.....N.........NC
HIV1_A_K89	3	24	NC...N..............NC
HIV1_A_KIG93	4	24	.C..N...N.........N..........NC
HIV1_A_SE8131	5	27	NC...N.........N....N.........NC
HIV1_A_SE6594	4	24	NC..N.....N........NC
HIV1_A_92UG037	4	27	.C..N...N.....N.........NC
HIV1_A_U455	5	23	.C.N...N...N..N.........NC.
C) HIV-1: M group			
HIV1_A_U455	5	23	.C.N...N...N..N.........NC.
HIV1_B_HXB2R	3	24	.C....N....N.............NC.
HIV1_C_ETH2220	3	20	.C....N.....N..........NC.
HIV1_D_84ZR085	5	25	NC.....N...NN...........NC.
HIV1_01_CM240	6	24	NC...N..N.....N......N......NC.
HIV1_F_VI850	4	25	NC.N..N.........NC.
HIV1_G_SE6165	7	29	NC.........NN...N....NN.........NC.
HIV1_H_90CF056	5	28	NC....NN..N...........NC.
HIV1_J_SE92809	5	24	NC.N.....N...N.........NC.
HIV1_K_MP535C	4	26	NC......N...N...........NC.
D) SIV_{CPZ}:			
CPZ_US	3	20	.C.....N....N......NC.
CPZ_GAB	4	22	.C...N.....N..N.........NC.
CPZ_CAM3	3	25	.C.....N.....N................NC.
CPZ_CAM5	3	23	.C....N......N...........NC.
CPZ_GAB2	4	22	.C...N...N..N.............NC.
CPZ_ANT	2	20	NC...N.........................C.
CPZ_TAN1	4	25	.C...N..N...........N.........NC.
E) HIV-2			
HIV2_A_GH1	5	24	.CN...NN...........N........NC.
HIV2_A_CBL21	3	22	NC....N................N..C.
HIV2_A_CAM3	2	26	.CN.................N....C.

Table I. Continued

Sequence name	No. *N*-linked glycosylation sites in V1	No. sites in full gp120	Spacing of sequons in V1 loops
HIV2_A_CBL23	3	24	.CN.....N.....N..C.
HIV2_A_CAM4	3	25	.C.....N.....N.....N....C.
HIV2_A_MDS	5	30	NC.....N.....N...N.....N.....C.
HIV2_A_FG	5	24	NC..N.....N....N.....NC.
HIV2_A_UC2	4	24	.C.....N..N.....N.....N....C.
HIV2_A_ALI	2	24	.C.N.....N....C.
HIV2_B_EHO	1	19	.CN.....C.
HIV2_B_D205	3	25	NC.....N.....N....C.
HIV2_B_UC1	3	23	.CN....N.....N....C.
HIV2_G_ABT96	2	23	.CN.....N....C.
F) SIV_{SMM}			
SMM_STM	2	22	.C..N.....N....C.
SMM_SL92B	2	23	.C..N.....N....C.
SMM_MM142	2	23	.CN.....N....C.
SMM_MNE027	2	22	.CN.....N..C.
SMM_MAC251	2	20	.CN.....N....C.
SMM_DeltaB670	2	23	.CN.....N....C.
SMM_F236	2	21	.CN.....N....C.
SMM_H9	2	21	.CN.....N....C.
G) SIV_{AGM}			
AGM_VER155	1	19	.C.....C.
AGM_VERAGM3	2	22	.C...N.....N.....C.
AGM_VER9063	2	22	.C...N....N.....C.
AGM_VERTYO	1	21	.C...N.....C.
AGM_SAB1C	2	20	.C.....N.....NC.
AGM_GRI677	1	21	.C...N.....C.
AGM_TAN1	1	20	NC.....C.

Only *N*-linked glycosylation sites (N) and Cys (C), which close the base of the V1 loop, are marked, all other amino acids are indicated by a period (.), to highlight the relative change of position of the *N*-linked sites and the length variation of the V1 loop. Parts C–G correspond to the sequences used in Figure 8A. Part A includes examples of V1 sequences taken from a single patient over time in a study of long-term survival (patient 20 from Shioda *et al.*, 1997), who was randomly selected from among longitudinal sequences sets spanning V1 available in the HIV database, and in our experience is representative. Part B includes seven randomly selected subtype A sequences from the HIV Los Alamos database alignment to represent intraclade diversity, respectively. Part C includes one randomly selected sequence from each clade to represent HIV-1 M group diversity. Part D includes available viral sequences isolated from chimpanzees, thought to be the source of the human HIV-1 epidemic. Part E lists viruses from HIV-2 clades A and B, and part F from sooty mangabey, which are thought to be the source of the HIV-1 epitdemic. Part G represents viral sequences taken from African green monkeys.

preserve the high-mannose forms. It has been suggested that the conservation of the high-mannose epitope for the 2G12 antibody might be related to the preservation of mannose structure that facilitates DC-SIGN interaction (Sanders *et al.*, 2002). *N*-linked glycosylation patterns in HIV Env are further complicated by the fact that different cell types, H9 and Chinese hamster ovary (Mizuochi *et al.*, 1988, 1990), and primary T cells and macrophages have different patterns of glycan modification (Liedtke *et al.*, 1997; Lin *et al.*, 2003; Willey *et al.*, 1996). DC-SIGN preferentially binds to HIV Env gp120 enriched for high-mannose oligosaccharides, which are typically produced

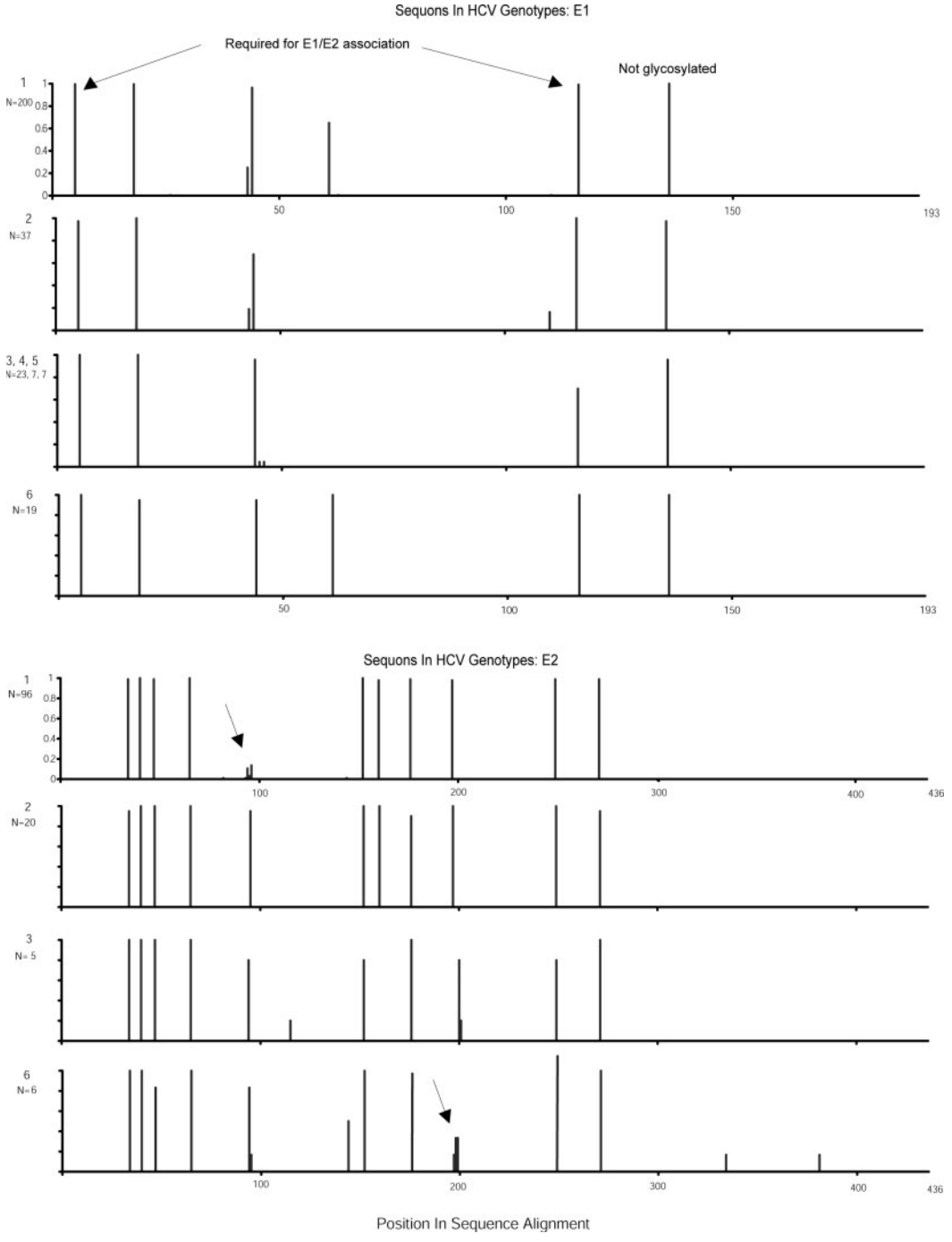

Fig. 9. *N*-linked glycosylation patterns in the E1 and E2 proteins of HCV. Sequon frequencies in HCV genotypes with adequate sampling are shown. The plot was nearly identical for E1 genotypes 3, 4, and 5, so only the genotype 3 plot is shown. Potential shifting sites are seen in E2 near position 90 in genotype 1, and near position 200 in genotype 6. The most C-terminal sequon in E1 is not glycosylated, and the sequon second from the C-terminal end and the most *N*-terminal together are critical for the formation of noncovalent E1E2 complexes (Meunier *et al*., 1999).

by peripheral blood mononuclear cells and T cells, compared to macrophage-produced gp120, which contains more complex carbohydrates (Lin *et al*., 2003). Also, macrophage-derived gp120 carbohydrates are modified by lactosaminoglycans, whereas peripheral blood mononuclear cells–derived gp120s are not. Interestingly, macrophage-derived virus tends to be more neutralization resistant (Willey *et al*., 1996). Bisecting *N*-glycans have been implicated in the suppression of natural killer (NK) cell–mediated responses (Yoshimura *et al*., 1996), the suppression of innate immune responses involving NK cells that may be responsible for initiating AIDS pathogenesis (Kottilil *et al*., 2003). Thus cell-specific glycosylation profiles may affect immune susceptibility, and AIDS progression may be related to the glycobiology of the virus (Clark *et al*., 1997).

Positions that accept complex carbohydrate additions may also be determined simply by being more exposed during passage through the Golgi apparatus, where carbohydrate modifications occur. Such exposure may be related to accessibility in the folded functional protein, and immune evasion and neutralizing antibody escape are likely to be the driving force for the variation in shifting sites. So the complex carbohydrate additions may occur in more exposed sites, and those sites, due to the exposure, may be under greater pressure to shift to escape from antibody recognition.

We did not observe variation in the net number of sequons in the variable loop domains of HIV-1 Env gp120 associated with any particular pattern of coreceptor usage, although a particular sequon in the V3 loop was highly conserved in R5 viruses but not in X4 or R5X4 viruses. Potential involvement of sequon changes in gp41 in cell tropism was also noted, as some sites are rarely found in NSI variants. Glycosylation in gp41 may be more influential than previously thought, given new evidence that additional loops of the transmembrane protein are located extracellularly (Cleveland *et al*., 2003). The role of alterations of specific sequons in HIV-1 Env associated with changes in coreceptor usage on a population basis can be subtle and may be context-specific. For example, in our analysis of one patient for changes associated with phenotypic variation (Figure 7), significant differences were found at positions that were different from those identified in the cross-sectional population analyses.

Glycosylation in the V1V2 region can be important for coreceptor usage (Ogert *et al*., 2001), and additional glycosylation sites in V1V2 may in some circumstances potentiate the use of CXCR4 (Pollakis *et al*., 2001). V2 elongation and sequon changes have been associated with slow disease progression (Shioda *et al*., 1997). The effects may be subtle; for example, removal of three sequons in V1 increased the affinity between gp120 and the CXCR4 receptor but did not alter the infectivity of the virus (Losman *et al*., 2001). Limited V1V2 length variation and sequon shifts in rapid progressors (Masciotra *et al*., 2002; Shioda *et al*., 1997) versus long-term survivors may be a consequence of a poor immune response resulting in weak selection pressure (Delwart *et al*., 1997; Wolinsky *et al*., 1996) and not related directly to the coreceptor usage. This is particularly plausible because changes in V1 V2 sequons often influence antigenic domains in other regions, for example, they can alter antibody recognition of both the V3 loop (Losman *et al*., 2001; Ly and Stamatatos, 2000; Ye *et al*., 2000) and CD4 binding site (Ly and Stamatatos, 2000). It is intriguing that the capacity for shifting sequons is found not only in HIV but also in two HCV E2 locations in two lineages in genotypes 1 and 6, suggesting they may give a selective advantage in rapidly evolving viruses.

Although the SIV_{SMM} and SIV_{AGM} viruses have some degree of shifting sequons in their natural hosts (particularly in the Env V4 region in SIV_{AGM}), the shifting Env sequon characteristics appear far less pronounced in these lineages than in HIV-1 and CPZ lineages (Figure 8). HIV-2, which stems from cross-species transmission of SIV_{SMM}, also has more extreme levels of shifting sites, suggesting that selective forces in the new host bring out the greater levels of position diversity seen in sequons. Neutralizing antibody responses to SIV_{SMM} and SIV_{AGM} infections in their natural hosts are present but may be relatively reduced (Fultz *et al*., 1990; Gicheru *et al*., 1999; Kaur *et al*., 1998; Norley *et al*., 1990). However, despite the lack of shifting sequons, both SIV_{SMM} and SIV_{AGM} diversify rapidly *in vivo* (Broussard *et al*., 2001). At least some of this diversification may be due to cytotoxic T lymphocyte escape (Kaur *et al*., 2001), which may not select for patterns of shifting sequons. Neutralizing activity of sera from HIV-infected individuals and SIV-infected primates generally lags behind, so serum from one time point can neutralize earlier but not contemporary virus (Albert *et al*., 1990; Arendrup *et al*., 1992; Bradney *et al*., 1999; Montefiori *et al*., 1991; Nyambi *et al*., 1997). Rapid cycles of response and escape may be related to the gain and loss of sequons (Richman *et al*., 2003; Wei *et al*., 2003). Thus it is possible that the strength and neutralizing antibody response in the host dictates the extent of the shifting antibody sites in different primate lentiviruses.

N-linked glycosylation sites are a critical component of the external proteins of primate lentiviruses, influenza, and hepatitis C viruses, and their modification can be important for evolution of the immune response. The gain or loss of such sites can play a key role in viral infectivity, antigen conformation, and immune escape. The mechanisms that generate shifting sites and tolerance of such shifting sequons in viral proteins provides a unique evolutionary avenue for immune evasion.

Materials and methods

A Web-based tool was developed for tracking and quickly assessing patterns in *N*-linked glycosylation sites in protein alignments (www.hiv.lanl.gov/content/hiv-db/GLYCOSITE/glycosite.html). This facilitates comparing glycosylation patterns by providing five different summaries:

i. A simple tally of the number of sequons in each protein in an alignment;
ii. Red highlighted *N*-linked glycosylation sites in the alignment, and a downloadable fasta-format text file that leaves the N in sequons uppercase while reducing all other amino acids to lowercase;
iii. Plots of the fraction of sequences in each alignment that carry an *N*-linked glycosylation site at each position;
iv. A summary of the average number of sites within a user-specified window size, and a break down for each sequence;
v. A list of sequons and their context in each sequence, providing the amino acid string of the sequons, as not all sequons have the same capacity to be glycosylated (for example, *N*-P-[ST] is not glycosylated), and such sites were not observed among the viral sequences studied here. We also include links to references defining glycosylation propensities of different local combinations of amino acids.

All features of the program were used to analyze the sequence sets included in this study, but most of the figures were made using feature iii.

The protein alignments used for this study were retrieved from the Los Alamos HIV sequence database (www.hiv.lanl.gov), Influenza database (www.flu.lanl.gov), and the Hepatitis C database (www.hcv.lanl.gov). All alignments used in this study are available on request. All hepatitis, SIV, and HIV alignments were restricted so that only sequence from a single individual was included. Thus the sample sets were not biased by including multiple sequences from one person or from closely related infections. The one exception is the sequon change analysis of single patient in Figure 7 (Hu *et al.*, 2000).

Acknowledgments

We thank Drs. John Moore and Julie Overbaugh for helpful discussions and Dennis Bruton for asking the questions lead to this study. This work was supported by the NIH-DOE interagency agreement for the HIV-1 sequence and immunology database project YI-AI-1500-03 (M.Z., B.G., B.F., C.K., B.K.), a Los Alamos National Laboratory research development award (B.K. and B.G.), and NIH-PHS grants AI-26503 (N.H.) and AI-34251 (N.H.). W.B. is supported by a University of Washington Center for AIDS Research fellowship (AI-27757).

Abbreviations

DC-SIGN, dendritic cell–specific HIV-1-binding protein; HA1, influenza A hemagglutinin 1; HCV, hepatitis C virus; HIV human immunodeficiency virus; NK: natural killer cell; NSI, nonsyncytium-inducing; CRF, circulating recombinant form; SI, syncytium-inducing; SIV, simian immunodeficiency virus.

References

Albert, J., Abrahamsson, B., Nagy, K., Aurelius, E., Gaines, H., Nystrom, G., and Fenyo, E.M. (1990) Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralization by autologous sera. *AIDS*, **4**, 107–112.

Alexander, S. and Elder, J.H. (1984) Carbohydrate dramatically influences immune reactivity of antisera to viral glycoprotein antigens. *Science*, **226**, 1328–1330.

Arendrup, M., Nielsen, C., Hansen, J.E., Pedersen, C., Mathiesen, L., and Nielsen, J.O. (1992) Autologous HIV-1 neutralizing antibodies: emergence of neutralization-resistant escape virus and subsequent development of escape virus neutralizing antibodies. *J. AIDS*, **5**, 303–307.

Bolmstedt, A., Sjolander, S., Hansen, J.E., Akerblom, L., Hemming, A., Hu, S.L., Morein, B., and Olofsson, S. (1996) Influence of *N*-linked glycans in V4–V5 region of human immunodeficiency virus type 1 glycoprotein gp160 on induction of a virus-neutralizing humoral response. *J. AIDS Hum. Retrovirol.*, **12**, 213–220.

Bosch, M.L., Andeweg, A.C., Schipper, R., and Kenter, M. (1994) Insertion of *N*-linked glycosylation sites in the variable regions of the human immunodeficiency virus type 1 surface glycoprotein through AAT triplet reiteration. *J. Virol.*, **68**, 7566–7569.

Botarelli, P., Houlden, B.A., Haigwood, N.L., Servis, C., Montagna, D., and Abrignani, S. (1991) *N*-glycosylation of HIV-gp120 may constrain recognition by T lymphocytes. *J. Immunol.*, **147**, 3128–3132.

Bradney, A.P., Scheer, S., Crawford, J.M., Buchbinder, S.P., and Montefiori, D.C. (1999) Neutralization escape in human immunodeficiency virus type 1–infected long-term nonprogressors. *J. Infect. Dis.*, **179**, 1264–1267.

Broussard, S.R., Staprans, S.I., White, R., Whitehead, E.M., Feinberg, M.B., and Allan, J.S. (2001) Simian immunodeficiency virus replicates to high levels in naturally infected African green monkeys without inducing immunologic or neurologic disease. *J. Virol.*, **75**, 2262–2275.

Bush, R.M., Fitch, W.M., Bender, C.A., and Cox, N.J. (1999) Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol. Biol. Evol.*, **16**, 1457–1465.

Chackerian, B., Morton, W.R., and Overbaugh, J. (1994) Persistence of simian immunodeficiency virus Mne variants upon transmission. *J. Virol.*, **68**, 4080–4085.

Chackerian, B., Rudensey, L.M., and Overbaugh, J. (1997) Specific *N*-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter recognition by neutralizing antibodies. *J. Virol.*, **71**, 7719–7727.

Cheng-Mayer, C., Brown, A., Harouse, J., Luciw, P.A., and Mayer, A.J. (1999) Selection for neutralization resistance of the simian/human immunodeficiency virus SHIVSF33A variant *in vivo* by virtue of sequence changes in the extracellular envelope glycoprotein that modify *N*-linked glycosylation. *J. Virol.*, **73**, 5294–5300.

Cho, M.W., Lee, M.K., Carney, M.C., Berson, J.F., Doms, R.W., and Martin, M.A. (1998) Identification of determinants on a dualtropic human immunodeficiency virus type 1 envelope glycoprotein that confer usage of CXCR4. *J. Virol.*, **72**, 2509–2515.

Clark, G.F., Dell, A., Morris, H.R., Patankar, M., Oehninger,S., and Seppala, M. (1997) Viewing AIDS from a glycobiological perspective: potential linkages to the human fetoembryonic defence system hypothesis. *Mol. Hum. Reprod.*, **3**, 5–13.

Cleveland, S.M., McLain, L., Cheung, L., Jones, T.D., Hollier, M., and Dimmock, N.J. (2003) A region of the C-terminal tail of the gp41 envelope glycoprotein of human immunodeficiency virus type 1 contains a neutralizing epitope: evidence for its exposure on the surface of the virion. *J. Gen. Virol.*, **84**, 591–602.

Davis, D., Stephens, D.M., Willers, C., and Lachmann, P.J. (1990) Glycosylation governs the binding of antipeptide antibodies to regions of hypervariable amino acid sequence within recombinant gp120 of human immunodeficiency virus type 1. *J. Gen. Virol.*, **71**(12), 2889–2898.

Deleersnyder, J., Pillez, A., Wychowski, C., Blight, K., Xu, J., Hahn, Y.S., Rice, C.M., and Dubuission, J. (1997) Formation of native hepatitis C virus glycoprotein complexes. *J. Virol.*, **71**, 697–704.

Delwart, E.L., Pan, H., Sheppard, H.W., Wolpert, D., Neumann, A.U., Korber, B., and Mullins, J.I. (1997) Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. *J. Virol.*, **71**, 7498–7508.

Dubuisson, J., Duvet, S., Meunier, J.C., De Beeck, O., Cacan, R., Wychowski, C., and Cocquerel, L. (2000) Glycosylation of the hepatitis C virus envelope protein E1 is dependent on the presence of a downstream sequence on the viral polyprotein. *J. Biol. Chem.*, **275**, 30605–30609.

Ferguson, N.M., Galvani, A.P., and Bush, R.M. (2003) Ecological and immunological determinants of influenza evolution. *Nature*, **422**, 428–433.

Ferris, R.L., Hall, C., Sipsas, N.V., Safrit, J.T., Trocha, A., Koup, R.A., Johnson, R.P., and Siliciano, R.F. (1999) Processing of HIV-1 envelope glycoprotein for class I–restricted recognition: dependence on TAP1/2 and mechanisms for cytosolic localization. *J. Immunol.*, **162**, 1324–1332.

Fournillier, A., Wychowski, C., Boucreux, D., Baumert, T.F., Meunier, J.C., Jacobs, D., Muguet, S., Depla, E., and Inchauspe, G. (2001) Induction of hepatitis C virus E1 envelope protein–specific immune response can be enhanced by mutation of *N*-glycosylation sites. *J. Virol.*, **75**, 12088–12097.

Fultz, P.N., Stricker, R.B., McClure, H.M., Anderson, D.C., Switzer, W.M., and Horaist, C. (1990) Humoral response to SIV/SMM infection in macaque and mangabey monkeys. *J. AIDS*, **3**, 319–329.

Gao, F., Morrison, S.G., Robertson, D.L., Thornton, C.L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B.,

von Briesen, H. and others. (1996) Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID networks for HIV isolation and characterization. *J. Virol.*, **70**, 1651–1667.

Gavel, Y. and von Heijne, G. (1990) Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.*, **3**, 433–442.

Gicheru, M.M., Otsyula, M., Spearman, P., Graham, B.S., Miller, C.J., Robinson, H.L., Haigwood, N.L., and Montefiori, D.C. (1999) Neutralizing antibody responses in Africa green monkeys naturally infected with simian immunodeficiency virus (SIVagm). *J. Med. Primatol.*, **28**, 97–104.

Gram, G.J., Bolmstead, A., Schonning, K., Biller, M., Hansen, J.E., Olofasson, S. (2002) Detection of orientation-specific anti-gp120 antibodies by a new *N*-glycanase protection assay. *APMIS*, **110**, 123–31.

Hebert, D.N., Zhang, J.X., Chen, W., Foellmer, B., and Helenius, A. (1997) The number and location of glycans on influenza hemagglutinin determine folding and association with calnexin and calreticulin. *J. Cell Bio.*, **139**, 613–623.

Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H., and Johnson, P.R. (1989) An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature*, **339**, 389–392.

Hoffman, T.L., Stephens, E.B., Narayan, O., and Doms, R.W. (1998) HIV type I envelope determinants for use of the CCR2b, CCR3, STRL33, and APJ coreceptors. *Proc. Natl Acad. Sci. USA*, **95**, 11360–11365.

Hu, Q.X., Barry, A.P., Wang, Z.X., Connolly, S.M., Peiper, S.C., and Greenberg, M.L. (2000) Evolution of the human immunodeficiency virus type 1 envelope during infection reveals molecular corollaries of specificity for coreceptor itilization and AIDS pathogenesis. *J. Virol.*, **74**, 11858–11872.

Kalish, M.L., Korber, B.T., Pillai, S., Robbins, K.E., Leo, Y.S., Saekhou, A., Verghese, I., Gerrish, P., Goh, C.L., Lupo, D., Tan, B.H., Brown, T.M., and Chan, R. (2002) The sequential introduction of HIV-1 subtype B and CRF01AE in Singapore by sexual transmission: accelerated V3 region evolution in a subpopulation of Asian CRF01 viruses. *Virology*, **304**, 311–329.

Kasturi, L., Eshleman, J.R., Wunner, W.H., and Shakin-Eshleman, S.H. (1995) The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence *N*-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.*, **270**, 14756–14761.

Kaur, A., Grant, R.M., Means, R.E., McClure, H., Feinberg, M., and Johnson, R.P. (1998) Diverse host responses and outcomes following simian immunodeficiency virus SIVmac239 infection in sooty mangabeys and Rhesus macaques. *J. Virol.*, **72**, 9597–9611.

Kaur, A., Alexander, L., Staprans, S.I., Denekamp, L., Hale, C.L., McClure, H.M., Feinberg, M.B., Desrosiers, R.C., and Johnson, R.P. (2001) Emergence of cytotoxic T lymphocyte escape mutations in nonpathogenic simian immunodeficiency virus infection. *Eur. J. Immunol.*, **31**, 3207–3217.

Kaverin, N.V., Rudneva, I.A., Ilyushina, N.A., Varich, N.L., Lipatov, A.S., Smirnov, Y.A., Govorkova, E.A., Gitelman, A.K., Lvov, D.K., and Webster, R.G. (2002) Structure of antigenic sites on the haemagglutinin molecule of H5 avian influenza virus and phenotypic variation of escape mutants. *J. Gen. Virol.*, **83**, 2497–2505.

Koito, A., Stamatatos, L., and Cheng-Mayer, C. (1995) Small amino acid sequence changes within the V2 domain can affect the function of a T-cell line-tropic human immunodeficiency virus type 1 envelope gp120. *Virology*, **206**, 878–884.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S., and Bhattacharya, T. (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science*, **288**, 1789–1796.

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001) Evolutionary and immunological implications of gontemporary HIV-1 variation. *Br. Med. Bull.*, **58**, 19–42.

Kottilil, S., Chun, T.W., Moir, S., Liu, S., McLaughlin, M., Hallahan, C.W., Maldarelli, F., Corey, L., and Fauci, A.S. (2003) Innate immunity in human immunodeficiency virus infection: effect of viremia on natural killer cell function. *J. Infect. Dis.*, **187**, 1038–1045.

Kuiken, C., Foley, B., Guzman E., and Korber, B. (1999) Determinants of HIV-1 protein evolution. In Crandall, K. (Ed.), *Molecular evolution of HIV*. Johns Hopkins University Press, Baltimore, MD.

Land, A. and Braakman, I. (2001) Folding of the human immunodeficiency virus type 1 envelope glycoprotein in the endoplasmic reticulum. *Biochimie*, **83**, 783–790.

Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N., and Gregory, T.J. (1990) Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.*, **265**, 10373–10382.

Liedtke, S., Geyer,R., and Geyer, H. (1997) Host-cell-specific glycosylation of HIV-2 envelope glycoprotein. *Glycoconj. J.*, **14**, 785–793.

Lin, G., Simmons, G., Pohlmann, S., Baribaud, F., Ni, H., Leslie, G.J., Haggarty, B.S., Bates, P., Weissman, D., Hoxie, J.A., and Doms, R.W. (2003) Differential *N*-linked glycosylation of human immunodeficiency virus and Ebola virus envelope glycoproteins modulates interactions with DC-SIGN and DC-SIGNR. *J. Virol.*, **77**, 1337–1346.

Losman, B., Bolmstedt, A., Schonning, K., Bjorndal, A., Westin, C., Fenyo, E.M., and Olofsson, S. (2001) Protection of neutralization epitopes in the V3 loop of oligomeric human immunodeficiency virus type 1 glycoprotein 120 by *N*-linked oligosaccharides in the V1 region. *AIDS Res. Hum. Retroviruses*, **17**, 1067–1076.

Ly, A. and Stamatatos, L. (2000) V2 loop glycosylation of the human immunodeficiency virus type 1 SF162 envelope facilitates interaction of this protein with CD4 and CCR5 receptors and protects the virus from neutralization by anti-V3 loop and anti-CD4 binding site antibodies. *J. Virol.*, **74**, 6769–6776.

Malenbaum, S.E., Yang, D., Cavacini, L., Posner, M., Robinson, J., and Cheng-Mayer, C. (2000) The *N*-terminal V3 loop glycan modulates the interaction of clade A and B human immunodeficiency virus type 1 envelopes with CD4 and chemokine receptors. *J. Virol.*, **74**, 11008–11016.

Marshall, R.D. (1974) The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem. Soc. Symp.*, 17–26.

Masciotra, S., Owen, S.M., Rudolph, D., Yang, C., Wang, B., Saksena, N., Spira, T., Dhawan, S., and Lal, R.B. (2002) Temporal relationship between V1V2 variation, macrophage replication, and coreceptor adaptation during HIV-1 disease progression. *AIDS*, **16**, 1887–1898.

Matrosovich, M., Zhou, N., Kawaoka, Y., and Webster, R. (1999) The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J. Virol.*, **73**, 1146–1155.

Matthews, T.J., Weinhold, K.J., Lyerly, H.K., Langlois, A.J., Wigzell, H., and Bolognesi, D.P. (1987) Interaction between the human T-cell lymphotropic virus type IIIB envelope glycoprotein gp120 and the surface antigen CD4: role of carbohydrate in binding and cell fusion. *Proc. Natl Acad. Sci. USA*, **84**, 5424–5428.

Mellquist, J.L., Kasturi, L., Spitalnik, S.L., and Shakin-Eshleman, S.H. (1998) The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of *N*-linked core glycosylation efficiency. *Biochemistry*, **37**, 6833–6837.

Meunier, J.C., Fournillier, A., Choukhi, A., Cahour, A., Cocquerel, L., Dubuisson, J., and Wychowski, C. (1999b) Analysis of the glycosylation sites of hepatitis C virus (HCV) glycoprotein E1 and the influence of E1 glycans on the formation of the HCV glycoprotein complex. *J. Gen. Virol.*, **80**(4), 887–896.

Mizuochi, T., Spellman, M.W., Larkin, M., Solomon, J., Basa, L.J., and Feizi, T. (1988) Carbohydrate structures of the human-immunodeficiency-virus (HIV) recombinant envelope glycoprotein gp120 produced in Chinese-hamster ovary cells. *Biochem. J.*, **254**, 599–603.

Mizuochi, T., Matthews, T.J., Kato, M., Hamako, J., Titani, K., Solomon, J., and Feizi, T. (1990) Diversity of oligosaccharide structures on the envelope gycoprotein gp 120 of human immunodeficiency virus 1 from the lymphoblastoid cell line H9. Presence of complex-type oligosaccharides with bisecting *N*-acetylglucosamine residues. *J. Biol. Chem.*, **265**, 8519–8524.

Montefiori, D.C., Zhou, I.Y., Barnes, B., Lake, D., Hersh, E.M., Masuho, Y., and Lefkowitz, L.B. Jr. (1991) Homotypic antibody

responses to fresh clinical isolates of human immunodeficiency virus. *Virology*, **182**, 635–643.

Moore, J.P. and Sodroski, J. (1996) Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J. Virol.*, **70**, 1863–1872.

Myers, G., MacInnes, K., and Korber, B. (1992) The emergence of simian/human immunodeficiency viruses. *AIDS Res. Hum. Retroviruses*, **8**, 373–386.

Norley, S.G., Kraus, G., Ennen, J., Bonilla, J., Konig, H., and Kurth, R. (1990) Immunological studies of the basis for the apathogenicity of simian immunodeficiency virus from African green monkeys. *Proc. Natl Acad. Sci. USA*, **87**, 9067–9071.

Norley, S., Beer, B., Holzammer, S., zur Megede, J., and Kurth, R. (1999) Why are the natural hosts of SIV resistant to AIDS? *Immunol. Lett.*, **66**, 47–52.

Nyambi, P.N., Lewi, P., Peeters, M., Janssens, W., Heyndrickx, L., Fransen, K., Andries, K., Vanden Haesevelde, M., Heeney, J., Piot, P., and van der Groen, G. (1997) Study of the dynamics of neutralization escape mutants in a chimpanzee naturally infected with the simian immunodeficiency virus SIVcpz-Ant. *J. Virol.*, **71**, 2320–2330.

Ogert, R.A., Lee, M.K., Ross, W., Buckler-White, A., Martin, M.A., and Cho, M.W. (2001) *N*-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. *J. Virol.*, **75**, 5998–6006.

Ohgimoto, S., Shioda, T., Mori, K., Nakayama, E.E., Hu, H., and Nagai, Y. (1998) Location-specific, unequal contribution of the N glycans in simian immunodeficiency virus gp120 to viral infectivity and removal of multiple glycans without disturbing infectivity. *J. Virol.*, **72**, 8365–8370.

Ohta, Y., Masuda, T., Tsujimoto, H., Ishikawa, K., Kodama, T., Morikawa, S., Nakai, M., Honjo, S., and Hayami, M. (1988) Isolation of simian immunodeficiency virus from African green monkeys and seroepidemiologic survey of the virus in various non-human primates. *Int. J. Cancer*, **41**, 115–122.

Overbaugh, J. and Rudensey, L.M. (1992) Alterations in potential sites for glycosylation predominate during evolution of the simian immunodeficiency virus envelope gene in macaques. *J. Virol.*, **66**, 5937–5948.

Overbaugh, J., Rudensey, L.M., Papenhausen, M.D., Benveniste, R.E., and Morton, W.R. (1991) Variation in simian immunodeficiency virus Env is confined to V1 and V4 during progression to simian AIDS. *J. Virol.*, **65**, 7025–7031.

Pollakis, G., S., Kang, Kliphuis, A., Chalaby, M.I., Goudsmit, J., and Paxton, W.A. (2001) *N*-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J. Biol. Chem.*, **276**, 13433–13441.

Polonoff, E., Machida, C.A., and Kabat, D. (1982) Glycosylation and intracellular transport of membrane glycoproteins encoded by murine leukemia viruses. Inhibition by amino acid analogues and by tunicamycin. *J. Biol. Chem.*, **257**, 14023–14028.

Ratner, L. (1992) Glucosidase inhibitors for treatment of HIV-1 infection. *AIDS Res. Hum. Retroviruses*, **8**, 165–173.

Rice, C.M. (1996) Flaviviridae: the viruses and their replication. In Fields, B.N., Knipe, D.M., and Howley, P.M. (Eds.), *Field's virology*, Lippincott-Raven Philidelphia, pp. 931–959.

Richman, D.D., Wrin, T., Little, S.J., and Petropoulos, C.J. (2003) Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl Acad. Sci. USA*, **100**, 4144–4149.

Robbins, K.E., Lemey, P., Pybus, O.G., Jaffe, H.W., Youngpairoj, A.S., Brown, T.M., Salemi, M., Vandamme, A.M., and Kalish, M.L. (2003) U.S. human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J. Virol.*, **77**, 6359–6366.

Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., and others. (2000) HIV-1 nomenclature proposal. *Science*, **288**, 55–56.

Sanders, R.W., Venturi, M., Schiffner, L., Kalyanaraman, R., Katinger, H., Lloyd, K.O., Kwong, P.D., and Moore, J.P. (2002) The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. *J. Virol.*, **76**, 7293–7305.

Sato, K., Okamoto, H., Aihara, S., Hoshi, Y., Tanaka, T., and Mishiro, S. (1993) Demonstration of sugar moiety on the surface of hepatitis C virions recovered from the circulation of infected humans. *Virology*, **196**, 354–357.

Scanlan, C.N., Pantophlet, R., Wormald, M.R., Ollmann, S.E., Stanfield, R., Wilson, I.A., Katinger, H., Dwek, R.A., Rudd, P.M., and Burton, D.R. (2002) The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1 → 2 mannose residues on the outer face of gp120. *J. Virol.*, **76**, 7306–7321.

Selby, M., Erickson, A., Dong, C., Cooper, S., Parham, P., Houghton, M., and Walker, C.M. (1999) Hepatitis C virus envelope glycoprotein E1 originates in the endoplasmic reticulum and requires cytoplasmic processing for presentation by class I MHC molecules. *J. Immunol.*, **162**, 669–676.

Shakin-Eshleman, S.H., Spitalnik, S.L., and Kasturi, L. (1996) The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of *N*-linked core-glycosylation efficiency. *J. Biol. Chem.*, **271**, 6363–6366.

Shioda, T., Oka, S., Xin, X., Liu, H., Harukuni, R., Kurotani, A., Fukushima, M., Hasan, M.K., Shiino, T., Takebe, Y. and others. (1997) *In vivo* sequence variability of human immunodeficiency virus type 1 envelope gp120: association of V2 extension with slow disease progression. *J. Virol.*, **71**, 4871–4881.

Si, Z., Cayabyab, M., and Sodroski, J. (2001) Envelope glycoprotein determinants of neutralization resistance in a simian-human immunodeficiency virus (SHIV-HXBc2P 3.2) derived by passage in monkeys. *J. Virol.*, **75**, 4208–4218.

Simmonds, P., Zhang, L.Q., McOmish, F., Balfe, P., Ludlam, C.A., and Brown, A.J. (1991) Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 Env sequences in plasma viral and lymphocyte-associated proviral populations *in vivo*: implications for models of HIV pathogenesis. *J. Virol.*, **65**, 6266–6276.

Skehel, J.J. and Wiley, D.C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.*, **69**, 531–569.

Slater-Handshy, T., Droll, D.A., Fan, X., Di Bisceglie, A.M., and Chambers, T.J. (2004) HCV E2 glycoprotein: mutagenesis of *N*-linked glycosylation sites and its effects on E2 expression and processing. *Virology*, **319**, 36–48.

Webby, R.J. and Webster, R.G. (2001) Emergence of influenza A viruses. *Phil. Trans. R. Soc. Lond B Biol. Sci.*, **356**, 1817–1828.

Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., Komarova, N.L., Nowak, M.A., Hahn, B.H., Kwong, P.D., Shaw, G. (2003) Antibody neutralization and escape by HIV-1. *Nature*, **422**, 307–312.

Wiley, D.C. and Skehel, J.J. (1987) The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.*, **56**, 365–394.

Willey, R.L., Shibata, R., Freed, E.O., Cho, M.W., and Martin, M.A. (1996) Differential glycosylation, virion incorporation, and sensitivity to neutralizing antibodies of human immunodeficiency virus type 1 envelope produced from infected primary T-lymphocyte and macrophage cultures. *J. Virol.*, **70**, 6431–6436.

Wills, C., Farmer, A., and Myers, G. (1996) Rapid sequon evolution in human immunodeficiency virus type 1 relative to human immunodeficiency virus type 2. *AIDS Res. Hum. Retroviruses*, **12**, 1383–1384.

Wolinsky, S.M., Wike, C.M., Korber, B.T., Hutto, C., Parks, W.P., Rosenblum, L.L., Kunstman, K.J., Furtado, M.R., and Munoz, J.L. (1992) Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science*, **255**, 1134–1137.

Wolinsky, S.M., Korber, B.T., Neumann, A.U., Daniels, M., Kunstman, K.J., Whetsell, A.J., Furtado, M.R., Cao, Y., Ho, D.D., and Safrit, J.T. (1996) Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science*, **272**, 537–542.

Wyatt, R. and Sodroski, J. (1998) The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science*, **280**, 1884–1888.

Ye, Y., Si, Z.H., Moore, J.P., and Sodroski, J. (2000) Association of structural changes in the V2 and V3 loops of the gp120 envelope glycoprotein with acquisition of neutralization resistance in a simian-human immunodeficiency virus passaged *in vivo*. *J. Virol.*, **74**, 11955–11962.

Yoshimura, M., Ihara, Y., Ohnishi, A., Ijuhin, N., Nishiura, T., Kanakura, Y., Matsuzawa, Y., and Taniguchi, N. (1996) Bisecting *N*-acetylglucosamine on K562 cells suppresses natural killer cytotoxicity and promotes spleen colonization. *Cancer Res.*, **56**, 412–418.

Yusim, K., Peeters, M., Pybus, O.G., Bhattacharya, T., Delaporte, E., Mulanga, C., Muldoon, M., Theiler, J., and Korber, B. (2001) Using human immunodeficiency virus type 1 sequences to infer historical features of the acquired immune deficiency syndrome epidemic and human immunodeficiency virus evolution. *Phil. Trans. R. Soc. Lond. B Biol. Sci.*, **356**, 855–866.

Zhu, X., Borchers, C., Bienstock, R.J., and Tomer, K.B. (2000) Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry*, **39**, 11194–11204.